

# Temporal Pattern of Online Communication Spike Trains in Spreading a Scientific Rumor: How Often, Who Interacts with Whom?

Ceyda Sanli<sup>1,\*</sup> and Renaud Lambiotte<sup>1</sup>

<sup>1</sup>Complexity and Networks, naXys, Department of Mathematics, University of Namur, 5000 Namur, Belgium

(Dated: September 2, 2015)

We study complex time series (spike trains) of online user communication while spreading messages about the discovery of the Higgs boson in Twitter. We focus on online social interactions among users such as retweet, mention, and reply, and construct different types of active (performing an action) and passive (receiving an action) spike trains for each user. The spike trains are analyzed by means of local variation, to quantify the temporal behavior of active and passive users, as a function of their activity and popularity. We show that the active spike trains are bursty, independently of their activation frequency. For passive spike trains, in contrast, the local variation of popular users presents uncorrelated (Poisson random) dynamics. We further characterize the correlations of the local variation in different interactions. We obtain high values of correlation, and thus consistent temporal behavior, between retweets and mentions, but only for popular users, indicating that creating online attention suggests an alignment in the dynamics of the two interactions.

PACS numbers: 89.65.-s, 05.45.Tp, 89.75.Fb

Keywords: Social dynamic behavior, Twitter social network, time series analysis, communication types in Twitter, classifying active and popular users, ranking activation and popularity

## I. INTRODUCTION

In recent years, online social media (OSM) have become a major communication channel, allowing users to share information in their social and professional circles, to discover relevant information pre-filtered by other users, and to chat with their acquaintances. In addition to their practical use for individuals, OSM have the advantage of generating a rich data set on collective social dynamics, as social relations among individuals, temporal properties of their interactions, and their contents are automatically stored. The study of these digital footprints has led to the emergence of computational social science, allowing to quantify at large-scales our political ideas and preferences [1], to discover roles in social network [2, 3], to predict our health [4] and personality [5], and to determine external effects on online behavior [6]. Importantly, in OSM, users are at the same time both actors and receivers and therefore the amplification of a trend originates from the interplay between influencing [7, 8] and being influenced [9–13].

A crucial aspect of OSM and more generally of human behavior is the underlying complex dynamics [14–17]. The time series of user activities, e.g. posting a tweet and replying to a message, are quite distinct from uncorrelated (Poisson random) dynamics in the presence of burstiness [18–20], temporal correlations [6, 21, 22], and non-stationarity of human daily rhythm [23, 24], which has significant implications. Diffusion on a temporal network cannot be accurately described by models on static networks and consequently the process presents non-Markovian features with strong influence on the time

required to explore the system [25, 26]. Furthermore, the dynamics drives a strong heterogeneity observed in user activity [27, 28] and user/content popularity [29–31]. Specifically, in Twitter, the heterogeneity in popularity has been observed and quantified in different ways by the size of retweet cascades, i.e. users re-transfer messages to their own followers with or without modifying them [32–36] or by the number of mentions of a user name, identified by the symbol @, in other people’s tweets [37].

In this paper, we focus on the dynamics of social interactions taking place when diffusing rumors about the discovery of the Higgs boson on July 2012 in Twitter [38]. Our main goal is to find connections between the statistical properties of user time series established on the same subject, e.g. the announcement of the discovery of the Higgs boson, and their activity and popularity. To this end, we analyze tweets including social interactions, such as retweets of a message (RT), mentions of a user name (@), and replies to a message (RE). For each type of the interactions, a user can either play an active, e.g. retweeting, or a passive, e.g. being retweeted, role. Therefore, we characterize each user by 8 time series: one active and one passive time series for each of the 3 types of interaction as well as for the aggregation of all interactions, as illustrated in Fig. 1. Active time series are denoted as WHO and passive time series are defined by WHOM. We then investigate whether the statistical properties of each signal is a good predictor for the activity and popularity of a user.

The following sections are organized as follows. In section II, we describe the data set and provide basic statistical properties of who and whom time series. In section III, we introduce a technique dedicated to the analysis of non-stationary time series, so-called local variation, originally established for neuron spike trains [39–42] and recently has been applied to hashtag spike trains in Twit-

---

\* cedaysan@gmail.com

ter [43, 44]. In section IV, we search for statistical relations between local variation and measures of popularity of a user. Finally, section V summarizes the key results and raises open questions.

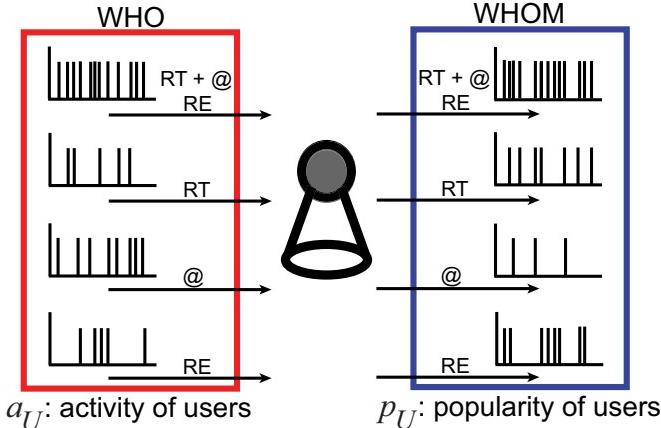


Figure 1. Illustration of communication in Twitter. Users in who interact in time with users in whom by retweeting (RT) the messages and mentioning (@) the user names of whom in a message and replying (RE) to the messages from whom. Quantifying temporal patterns in time series of who users with various ranges of the activity of users  $a_U$  and of whom users by increasing the popularity of users  $p_U$  is the main scope of this paper.

## II. ACTIVITY AND POPULARITY OF USERS

Our aim is to examine the dynamics of user communication in Twitter. We investigate how frequently Twitter users talk to each other on a certain topic, e.g. the discovery of the Higgs boson, and identify how complex dynamic patterns of the communication evolve in time. To this end, we focus on the 3 different types of interaction between users, retweet (RT), mention (@), and reply (RE). Twitter users can adopt a tweet of someone and use it again in their own tweet blog by RT or contact to other users directly by typing user names in a message called @ or simply RE to any tweets, e.g. regular tweets, retweets, and tweets/retweets including @s. Typically, @s and REs are associated to personal interactions between users, whereas RTs are responsible for large-scale information diffusion in the social network and present cascades. Here, we count all types of interaction as a part of complex information diffusion in Twitter.

Interactions in Twitter are performed between at least two users (for instance, a user can mention several other users in a single tweet). Each action is directed and characterized by its timestamp. The users performing the action play active roles (who users) and the users receiving their attention play a passive role (whom users). We construct active and passive RT, @, and RE spike trains for each user.

**Data Set.** As a test bed, we consider the publicly available Higgs Twitter data set [38, 45], first collected to track the spread of the rumor on the discovery of the Higgs boson via RT, @ or RE. The data set is composed of tweets containing one of the following keywords or hashtags related to the discovery of the Higgs boson, “lhc”, “cern”, “boson”, and “higgs”. The start date is the 1st July 2012, 00:00 am and the final date is the 7th July 2012, 11:59 pm, which covers the announcement date of the discovery, the 4th July 2012, 08:00 am. All dates and timestamps in the data are converted to the Greenwich mean time. Detailed information on the data collection procedure and basic statistics can be found in Ref. [38].

In total, the data is composed of 456,631 users (nodes) and 563,069 interactions. Among those, we detect 354,930 RT, 171,237 @, and 36,902 RE, which shows that RT is more popular than the other communication channels. For RT interactions, we find 228,560 who and 41,400 whom users. These numbers are smaller for @, e.g. 102,802 who and 31,477 whom users, and even smaller for RE, with 27,227 who and 18,578 whom users. In each case, whom is much lower than who, as expected because a small number of users tend to attract a large fraction of attention in both friendship [46, 47] and online social [48–52] networks. This observation is confirmed in Fig. 2, where we present the Zipf plots associated to each interaction, clearly showing a strong heterogeneity in the system. For who users, the frequency of the user communication  $f_U$  ranks how active users are and measures the activity of users  $a_U$ , on the other hand, for whom users,  $f_U$  quantifies how often the users or their tweets are addressed and so gives the popularity of users  $p_U$ .

## III. LOCAL VARIATION OF WHO AND WHOM

**Communication Spike Trains.** We extract salient temporal patterns of the user communication time series. We evaluate each directed interaction (RT, @, and RE) of the users in the pool of who with any users in the whom class, shown in Fig. 1. We don't check whether the whom users participate the conversation in a later stage and only construct independent time series of the who and whom users. The elements of the time series are the timestamps of the data [38, 45] providing us the exact time in second of the interaction and the user name or ID of the corresponding who and whom users. Ordering the timestamps from the earliest to the latest, we generate spike trains carrying full story of the communication of each user. The resultant user communication spike trains are grouped in eight: For each who and whom, the spike trains of all interactions together (i) and the spike trains of filtered timestamps of RT (ii), @ (iii), and RE (iv).

**Local Variation.** A standard way of investigating the dynamics of human communication is to examine the statistics of the inter-event spike intervals such as its probability distribution [14], short-range memory coefficient and burstiness parameter [15] or Fano factor. How-

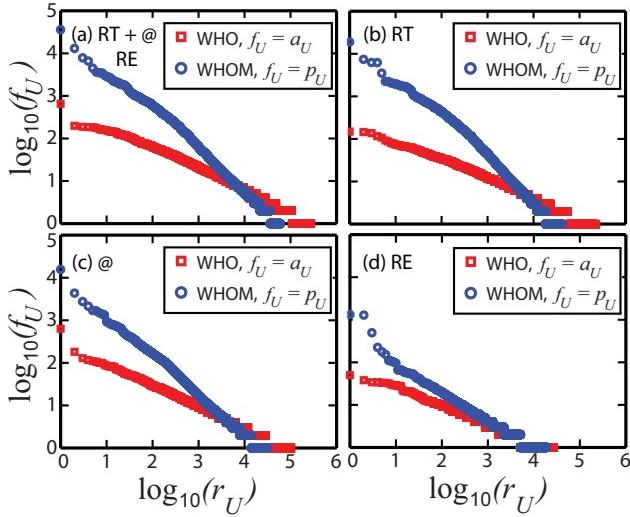


Figure 2. How often who users communicate with whom users. Zipf plots describe heterogeneities of users in the types of Twitter interaction, e.g. (a) all of retweet, RT, mention, @, and reply, RE, (b) only RT, (c) only @, and finally (d) only RE. The frequency of the communication  $f_U$  is measured in two-fold: The activity of who users (red squares)  $a_U$  and the popularity of whom users (blue circles)  $p_U$ . The  $x$ -axis ranks the users  $r_U$  from high  $f_U$  to low values. Each plot indicates that Twitter users, who users, more likely contact to someone observed in smoother decays of who users, however only few whom users are addressed and become popular in these communications.

ever, recent works have showed that further detail analysis is required to resolve temporal correlations [31, 32], bursts [19–22], and cascading [53] driven by circadian rhythm [23, 24], complex decision-making of individuals [3, 27, 54], and external factors [6] such as the announcement of discoveries, as considered in the current data [38].

To uncover the dynamics of the communication spike trains elaborately, we apply the local variation  $L_V$  originally defined to characterize non-stationary neuron spike trains [39–42] and very recently has been used to analyze hashtag spike trains [43, 44]. Unlike to the memory coefficient and burstiness parameter [15],  $L_V$  provides a local temporal measurement, e.g. at  $\tau_i$  of a successive time sequence of a spike train  $\dots, \tau_{i-1}, \tau_i, \tau_{i+1}, \dots$ , and so compares temporal variations with their local rates [41]

$$L_V = \frac{3}{N-2} \sum_{i=2}^{N-1} \left( \frac{(\tau_{i+1} - \tau_i) - (\tau_i - \tau_{i-1})}{(\tau_{i+1} - \tau_i) + (\tau_i - \tau_{i-1})} \right)^2 \quad (1)$$

where  $N$  is the total number of spikes. Eq. 1 also takes the form [41]

$$L_V = \frac{3}{N-2} \sum_{i=2}^{N-1} \left( \frac{\Delta\tau_{i+1} - \Delta\tau_i}{\Delta\tau_{i+1} + \Delta\tau_i} \right)^2 \quad (2)$$

Here,  $\Delta\tau_{i+1} = \tau_{i+1} - \tau_i$  quantifying the forward delays and  $\Delta\tau_i = \tau_i - \tau_{i-1}$  representing the backward waiting

times for an event at  $\tau_i$ . Importantly, the denominator normalizes the quantity such as to account for local variations of the rate at which events take place. By definition,  $L_V$  takes values in the interval (0:3) [43]. It has been shown that  $L_V$  classifies the salient dynamic patterns successfully [39, 40, 42–44]. Following the analysis of Gamma processes [39, 40, 43] conventionally applied to model inter-event intervals and the neuron spike analysis [42], while  $L_V = 1$  for uncorrelated (Poisson random) irregular spike trains,  $L_V \approx 3$  proves that bursts dominate the spike trains and the presence of highly regular patterns in the trains gives  $L_V \approx 0$ .

We now investigate the  $L_V$  analysis on the user communication spike trains. Eq. 2 is performed through the spike trains with removing multiple spikes taking place within one second. Such events are rare and their impact on the value of  $L_V$  has been shown to be limited [43]. Fig. 3 describes the distribution of  $L_V$ ,  $P(L_V)$  of full spike trains all together with RT, @, and RE for the who (a, b) and whom (c, d) users. Grouping  $L_V$  based on the frequency  $f_U$ , e.g. the activity of the who users  $a_U$  and the popularity of the whom users  $p_U$ , we examine the temporal patterns of the trains in different classes of  $a_U$  and  $p_U$ . For the real data in (a, c), in Fig. 3(a),  $L_V$  is always larger than 1 in any values of  $a_U$ , suggesting that all who users contact to the whom users in bursty communications. However, in Fig. 3(c), we observe distinct behavior of the whom users and bursts present only for low  $p_U$ . By increasing  $p_U$ ,  $L_V \approx 1$  indicating that there is no temporal correlations among the who users referring the whom users and  $L_V$  is slightly smaller than 1 for the most popular users, indicating a tendency towards regularity in the time series, as also observed for the hashtag spike trains [43]. These observations are significantly different for artificial spike trains constructed by randomly permuting the real full spike train and so expected to generate non-stationary Poisson processes. Therefore, all distributions are centered around 1 in this case, independently of  $a_U$  and  $p_U$ , as shown in Figs. 3(b, d). The randomization and obtaining a null set follow the same procedure explained in detail in Ref. [43].

Even though Fig. 3 represents  $P(L_V)$  of full spike trains, i.e. all interactions together,  $P(L_V)$  of individual RT, @, and RE communication spike trains describes very similar temporal behavior for both the who and whom users. Fig. 4 summarizes the detail of  $P(L_V)$ , the mean of  $L_V$ ,  $\mu(L_V)$  with the corresponding standard deviations  $\sigma(L_V)$  as error bars, comparatively. The results highlight that to classify the communication temporal patterns neither the position of the users, whether active or passive, nor the types of the interaction, but the frequency of the communication  $f_U$  such as  $a_U$  and  $p_U$  plays a major role. All Figs. 4(a-d), we observe three regions: Bursts in low  $f_U$ ,  $\log_{10}\langle f_U \rangle < 2.5$ , irregular uncorrelated (Poisson random) dynamics in moderate and high  $f_U$ ,  $\log_{10}\langle f_U \rangle \approx 2.5\text{-}3$ , and regular patterns in very high  $f_U$ ,  $\log_{10}\langle f_U \rangle > 3$ . This conclusion supports the importance of frequency so time parameter overall human

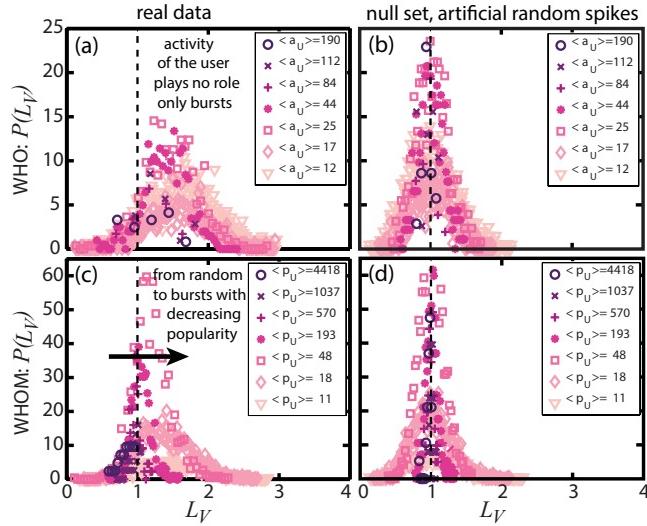


Figure 3. Probability density function of the local variation  $L_V$ ,  $P(L_V)$  of who (a, b) and whom (c, d) users in various ranges of the two communication frequencies, e.g.  $a_U$  and  $p_U$ . (a, c) describe the results of the real data. When we only observe bursty communication patterns in who users independent of the average user activity frequency  $\langle a_U \rangle$  in (a), significant variations in  $L_V$  by increasing the average user popularity  $\langle p_U \rangle$  are clear in (c). The results prove that popular users in Twitter are addressed randomly in time and slightly more regular patterns observed in the most popular users. On the other hand, (b, d) present the statistics of artificially generated random spikes serving as a null model and all frequency ranges give the distributions around 1, as expected for temporally uncorrelated signals.

behavior [14, 16].

We now perform more detail comparison in Fig. 5, how  $L_V$  of different interactions in the same frequency range varies from each other. To this end, we calculate the standard  $z$ -values in two ways. First, to compare  $L_V$  of the full spike trains with  $L_V$  of only RT and also with  $L_V$  of only @ spike trains,  $L_V^{\text{RT}}$  and  $L_V^{\text{@}}$ , respectively, we introduce

$$z(f_U) = \frac{\mu(L_V^k) - \mu_0(L_V)}{\sigma(L_V^k)/\sqrt{f_U^k}} \quad (3)$$

Here,  $k$  in superscripts labels the interaction, e.g. either RT or @. Precisely,  $L_V^k$  is determined based on a filtered spike train composed of the user timestamps of either RT or @, as already used in Fig. 4(b-c). In addition,  $\mu^k$  is the mean of  $L_V^k$ , also presented in Fig. 4(b-c), and  $\mu_0$  is the mean  $L_V$  of the full spike train, given in Fig. 4(a).

In Fig. 5, black squares show  $z$ -values of RT and black circles describe  $z$ -values of @. For the who users in Fig. 5(a) where  $L_V$  only presents bursty patterns (orange shaded area) and low  $a_U$ , we have small  $z$ -values proving the agreement of the temporal patterns suggested by  $L_V$  in the same  $a_U$ . However, for the whom users in Fig. 5(b) where we have rich values of  $p_U$  compared to the values of  $a_U$ , while  $z$ -values are small in bursty patterns (low

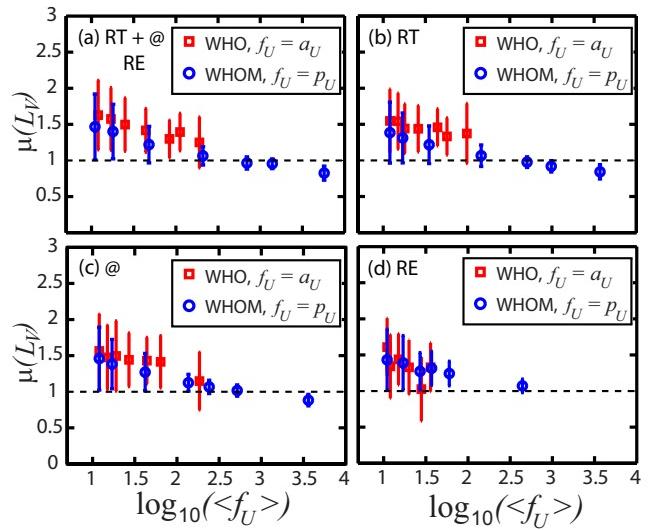


Figure 4. Mean  $\mu$  of the local variation  $L_V$  of the user communication spike trains versus the logarithmic average frequency  $\log_{10}\langle f_U \rangle$ . The results of who users are represented by red squares and blue circles describe that of whom users. Types of the interaction are investigated in detail: (a) All communications of retweet, RT, mention, @, and reply, RE. (b) Only RT. (c) Only @. (d) Only RE. Independent of the types of the interaction, the frequency of communication, e.g. the activity of users  $a_U$  and the popularity of users  $p_U$ , designs overall communication patterns. While low  $f_U$  gives bursty patterns with  $L_V > 1$ , moderate  $f_U$  indicates irregular uncorrelated (Poisson random) signals, e.g.  $L_V \approx 1$ . For all high  $f_U$ ,  $L_V < 1$  presenting the regularity of the communications. The error bars show the corresponding standard variations.

$p_U$ , orange area) as also observed in the who users and in regular patterns (high  $p_U$ , yellow area), larger  $z$ -@ value (the black circle) is calculated in uncorrelated Poisson dynamics (moderate  $p_U$ , purple area). The disagreement of  $L_V$  with large  $z$ -@ indicates that even though  $L_V \approx 1$  in this region the results of @ are quite sensitive in the same  $p_U$ , which is not observed in  $z$ -RT (the black square).

Furthermore, we repeat the analysis across communication channels by comparing temporal patterns of RT and @ as follows

$$z(f_U) = \frac{\mu(L_V^{\text{@}}) - \mu_0(L_V^{\text{RT}})}{\sigma(L_V^{\text{@}})/\sqrt{f_U^{\text{@}}}} \quad (4)$$

The corresponding  $z$ -values,  $z$ -@RT are presented in green diamonds in Fig. 5. Comparing to the previous  $z$ -RT and  $z$ -@, we now obtain lower values for the who users [Fig. 5(a)] showing a good agreement between RT and @ patterns. Moreover, we have very similar trend for the whom users [Fig. 5(b)] as before in orange and yellow areas and large fluctuations are observed only in purple area.

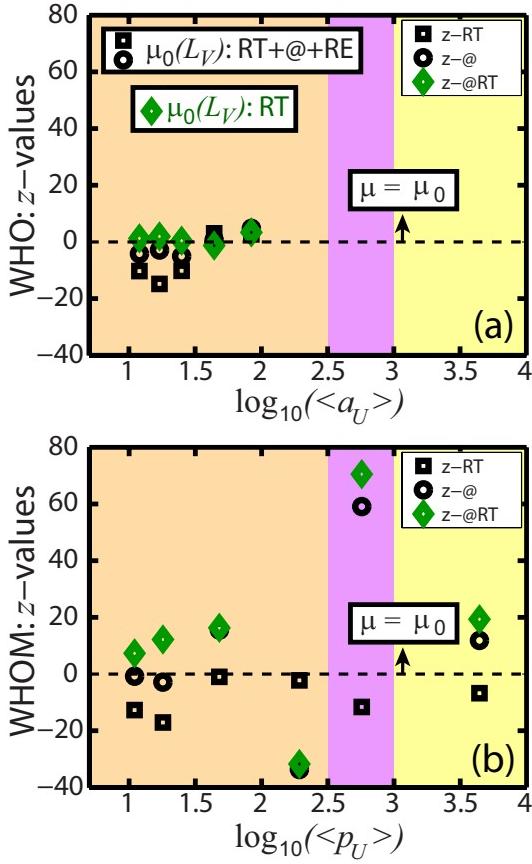


Figure 5. Detail comparison between the temporal patterns of different communications in each frequency range. While  $x$ -axis is the logarithmic average of frequency, e.g. (a)  $\log_{10}\langle a_U \rangle$  for the who users and (b)  $\log_{10}\langle p_U \rangle$  for the whom users,  $y$ -axis provides the calculation of three  $z$ -values (i)  $z$ -RT, the comparison of  $L_V$  of the full spike train with  $L_V$  of RT, in black squares, (ii)  $z$ -@, the same with  $L_V$  of @, presented in black circles, and (iii)  $z$ -@RT, the comparison between  $L_V$  of @ and RT, shown in green diamonds. All  $z$ -values are consistent with each other such that except moderate frequency range in (b), e.g.  $z$ -@ and  $z$ -@RT, we observe small  $z$  concluding that the temporal patterns in the similar frequency ranges are in a good agreement. Three distinct regions are colored due to the discovered patterns in calculating  $L_V$  in Fig. 4. Orange shaded area describes the ranges of the bursty patterns ( $a_U$  and low  $p_U$ ), purple area is for the irregular uncorrelated -Poisson random- patterns (moderate  $p_U$ ), and yellow area covers the regular patterns (high  $p_U$ ).

#### IV. CORRELATION OF $L_V$ IN USER COMMUNICATION HABITS

In this final section, our interest turns into building new measures to quantify how the local variation  $L_V$  fluctuates inside different classes of the frequency,  $f_U$ . What extend temporal communication habits of two users in the same  $f_U$  ranges are dependent on each other is the first question we address. Second, we examine whether the temporal patterns of the interactions are consistent

with each other for the same users and how the metric varies with increasing  $f_U$ .

We consider  $r_{ij}^{kk'}(f_U)$ , the Pearson correlation coefficient of  $L_V$  of two different users selected independently from the same  $f_U$  classes

$$r_{ij}^{kk'}(f_U) = \frac{\sum_{i,j=1,i \neq j}^{N_U} [L_{V_i}^k - \mu(L_{V_i}^k)][L_{V_j}^{k'} - \mu(L_{V_j}^{k'})]}{\sigma(L_{V_i}^k)\sigma(L_{V_j}^{k'})} \quad (5)$$

where  $\sigma(L_{V_i}^k) = \sqrt{\sum_{i=1}^{N_U} [L_{V_i}^k - \mu(L_{V_i}^k)]^2}$ . Here,  $L_{V_i}$  and  $L_{V_j}$

are the local variations of user  $i$  and  $j$ , respectively,  $\mu$ 's are the corresponding mean values, and  $N_U$  is the total number of users. Moreover,  $k$  and  $k'$  represent all permutations among the full, RT, and @ spike trains. Furthermore,  $r_{ij}^{kk'}(f_U)$  is evaluated for the who and whom users, separately. Therefore,  $i$  and  $j$  are different users, but from the same (who/whom) pool and in the same frequency classes of  $a_U$  and  $p_U$ , as grouped in Fig. 3. Note that before performing Eq. 5, the corresponding  $L_V$ 's in the same  $f_U$  class are ordered from the highest to the smallest (or vice versa) not to deform  $r_{ij}^{kk'}(f_U)$  artificially due to the random selection.

Fig. 6 presents the results of  $r_{ij}^{kk'}(f_U)$  for the who users in (a, b) and the whom users in (c, d). Similar to  $z$ -values performed in the previous Section, we suggest three correlation coefficients: Red (left) triangles describe  $r_{ij}^{\text{full,RT}}$ , blue (right) triangles are for  $r_{ij}^{\text{full,@}}$ , and black and green diamonds show the values of  $r_{ij}^{\text{RT,@}}$ . The average frequency of the users  $\langle f_U \rangle$  in the same class is similar but not equal and that is why Figs. 6(b, d) are plotted with respect to both the mean frequencies of RT and @, e.g. the average activity  $\langle a_U \rangle$  and popularity  $\langle p_U \rangle$  of RT and @. All correlations are above 0.85 proving the high dependency of the communication patterns of the users in the same  $\langle f_U \rangle$ , independent of the types of the interaction.

We now consider Eq. 5 with imposing the same user and repeat the procedure above for the correlation coefficient

$$r_i^{kk'}(f_U) = \frac{\sum_i^{N_U} [L_{V_i}^k - \mu(L_{V_i}^k)][L_{V_i}^{k'} - \mu(L_{V_i}^{k'})]}{\sigma(L_{V_i}^k)\sigma(L_{V_i}^{k'})} \quad (6)$$

Fig. 7 summarizes the results of Eq. 6. While Figs. 7(a, c) are in parallel with that of Fig. 6 with slightly lower correlations for @ (blue right triangles), distinct behavior is observed in Figs. 7(b, d). Low correlations in Fig. 7(b) indicate that the same who users present different temporal behavior in RT and @. On the other hand, Fig. 7(d) shows an interesting temporal habit of the whom users. Having no remarkable dependency captured in low popular users, we show that the correlation increases with  $\langle p_U \rangle$  describing that the popular users are addressed in RT and @ in a temporarily similar procedure.

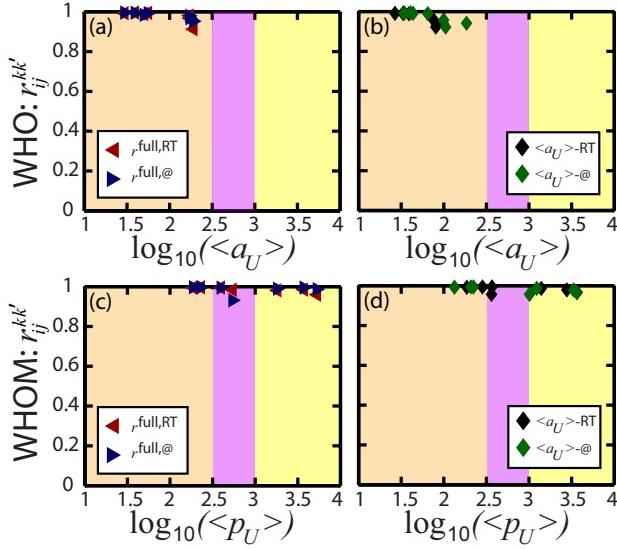


Figure 6. Linear correlations of  $L_V$  of user pairs: The standard Pearson correlation coefficient quantifies the dependency on the temporal communication habits of two different users independently chosen from the same frequency classes, as introduced in Fig. 3. The coefficient covers 3 potential relations in the communication interactions, e.g. full and RT spike trains, red (left) triangles, full and @, blue (right) triangles, and finally RT and @, black and green diamonds. These 3 coefficients are calculated for who (a, b) and whom users (c, d), separately. 6 coefficients in total prove that the temporal patterns present high consistency in each average frequency classes, the activity  $\langle a_U \rangle$  and the popularity  $\langle p_U \rangle$ . In (b, d), the corresponding coefficients are described with the sensitivity of the frequency classes since the average frequency in the class of RT is so similar, but not exactly equal to that of @. The colored areas are as defined in Fig. 5 and characterize the three main regions of the temporal patterns of the individual user spike trains, e.g. bursts (orange), irregular random (purple), and regular patterns (yellow).

### A. Nomenclature

- OSM: Online Social Media,
- @: Mention a user name in a tweet message,
- RE: Reply to a tweet or retweet message,
- RT: Retweet, share a message of other users in her/his own tweet blog,
- WHO: Twitter users starting an interaction via @ or RE or RT with any other users,
- WHOM: Twitter users addressed by the who users such that their message is retweeted or user name is mentioned in a message by the who users or they get a reply from the who users.

Any relation between who and whom users such as the following-follower is not imposed.

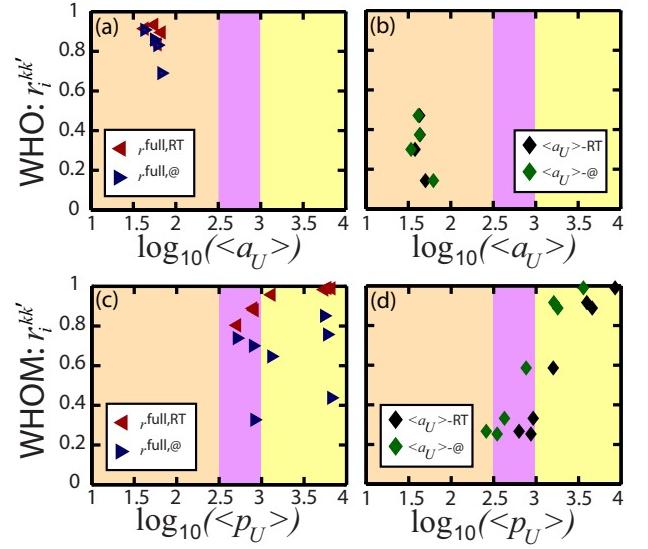


Figure 7. Linear correlations of  $L_V$  of the same users: The procedure and representation of the coefficients follow the same procedure as introduced in Fig. 6. However, we now impose the same users in the same frequency classes. Even though (a, c) present the agreement in the temporal patterns of full and RT spike trains of the same users, with high correlation coefficients in almost all frequency ranges, (b) indicates lower consistency between RT and @ spike trains during entire activity  $\langle a_U \rangle$  and (d) provides a significant result. While less temporal coherence is observed between RT and @ spike trains in low popularity  $\langle p_U \rangle$ , the correlation drastically increases with  $\langle p_U \rangle$ .

## V. DISCUSSION

In this paper, our interest is to quantify online user communication in Twitter. To reduce the complexity in the communication, the data studied here consider only a unique subject which users talk about such as the discovery of the Higgs boson on July 4, 2012 within a restricted time window, e.g. 6 days [38]. The main aim is to extract salient temporal patterns of communication in various types of interaction observed in Twitter such as retweet (RT), mention (@), and reply (RE). Adopting the technique so-called local variation  $L_V$  originally introduced for neuron spike trains [39–42] and recently has applied to hashtag spike trains in Twitter [43, 44], we perform detail analysis on user communication spike trains. Showing strong influences of the frequency of the hashtag spike trains on the resultant temporal patterns in the earlier work [43, 44], in parallel we here examine the differences in the patterns induced by the frequency of the user communication spike trains,  $f_U$ .

We investigate user communication spike trains in two categorizations, first set of users are the active ones, who users, and the other set is composed of the passive users, whom users, in the communication. For the who users,  $f_U$  simply gives what extend the users contact to the whom users and so it is the activity of the who users,

$a_U$ . On the other hand, for the whom users, the generated spike trains present how often the who users refer the messages or the user names of the whom users and therefore,  $f_U$  is the popularity of the whom users,  $p_U$ . Providing comparative statistics on  $L_V$  of who and whom with increasing  $a_U$  and  $p_U$ , respectively, we observe quite distinct temporal behavior of online users. First, we observe an asymmetry between active and passive interactions, as only the former give rise to hubs, with few users attracting a large share of the attention. Moreover, who users constantly present bursty patterns,  $L_V > 1$  for all values of  $a_U$ , whereas whom users demonstrate various dynamic behavior patterns, depending on their popularity: The least popular users with low  $p_U$  experience bursty time series, popular users with moderate and high  $p_U$  are contacted by temporarily uncorrelated whom users and so show Poisson random spike trains  $L_V \approx 1$ , and the most popular users with the maximum  $p_U$  are referred regularly in time, e.g.  $L_V < 1$ .

These scenarios are independent of both the position of the users, e.g. who or whom, and the preferred interactions, e.g. whether RT or @, suggesting that the frequency of the communication dominates to design social dynamic behavior. This conclusion is also supported by the high correlation coefficient of  $L_V$  on the user pairs in the same frequency classes. Furthermore, the linear correlation of  $L_V$  on the same users reveals interesting patterns. There, we observe that only popular users have similar dynamic behavior in both RT and @, which confirms that both metrics are complementary to characterize the influence of users.

The analysis could be specified by integrating the communication spike trains with the following-follower relation in Twitter, and focusing on the who and whom trains of connected users. An important concern is the limited time period of the data which the collection started 3 days before the announcement of the discovery and continued until 3 days after this date. Yet, it has been shown that the dynamics of the communication is drastically different before/after and during the announcement [38], and this variation could be investigated in our analysis. Our study shares the similar aims of the other research on online user behavior and the influence of the frequency

in online platforms such as Flickr, Delicious and StumbleUpon, which user profiles have been included in the analysis [47]. This understanding could be also applied to our analogy with considering further details in the data.

## A. Data Sharing

The full data studied in this paper has open access [38, 45].

## DISCLOSURE/CONFLICT-OF-INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: CS. Performed the experiments: CS. Analyzed the data: CS. Contributed reagents/materials/analysis tools: RL CS. Wrote the paper: CS RL.

## ACKNOWLEDGMENTS

C. Sanli acknowledges supports from the European Union 7th Framework OptimizR Project and FNRS (le Fonds de la Recherche Scientifique, Wallonie, Belgium). This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office.

*Funding.* The EU 7th Framework OptimizR Project: 48909A2 CE OPTIMIZR (Grant holder: RL, Funding receiver: CS - <http://optimizr.eu/>) and F.N.R.S MIS F4527.12 48888F3 (Grant holder: RL, Funding receiver: CS - <http://www.fnrs.be/>).

- 
- [1] A. Abisheva, V. R. K. Garimella, D. Garcia, and I. Weber, Proceedings of the 7th ACM International Conference on Web Search and Data Mining , 593 (2014).
  - [2] S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts, Proceedings of the 20th International Conference on World Wide Web , 705 (2011).
  - [3] Q. Xuan, M. Gharehyazie, P. Devanbu, and V. Filkov, Proceedings of the 2012 International Conference on Social Informatics , 78 (2012).
  - [4] M. D. Choudhury, Gamon, Counts, and Horvitz, Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media , 128 (2013).
  - [5] R. Lambiotte and M. Kosinski, Proceedings of the IEEE **102**, 1934 (2014).
  - [6] J. Mathiesen, L. Angheluta, P. T. H. Ahlgren, and M. H. Jensen, Proceedings of the National Academy of Sciences **110**, 17259 (2013).
  - [7] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, Proceedings of the Fourth ACM International Conference on Web Search and Data Mining , 65 (2011).
  - [8] S. Gonzalez-Bailon, J. Borge-Holthoefer, and Y. Moreno, American Behavioral Scientist **57**, 943 (2013).
  - [9] E. Bakshy, B. Karrer, and L. A. Adamic, Proceedings of the 10th ACM Conference on Electronic Commerce ,

- 325 (2009).
- [10] Y.-R. Lin, J. P. Bagrow, and D. Lazer, Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media , 193 (2011).
- [11] K. H. Kwon, M. A. Stefanone, and G. A. Barnett, American Behavioral Scientist **58**, 1345 (2014).
- [12] J. Zhang, J. Tang, J. Li, Y. Liu, and C. Xing, ACM Trans. Knowl. Discov. Data **9**, 25:1 (2015).
- [13] K. Lerman, X. Yan, and X.-Z. Wu, ArXiv e-prints (2015), arXiv:1506.03022.
- [14] A.-L. Barabasi, Nature **435**, 207 (2005).
- [15] K.-I. Goh and A.-L. Barabasi, EPL (Europhysics Letters) **81**, 48002 (2008).
- [16] G. Miritello, R. Lara, and E. Moro, Temporal Networks , 175 (2013).
- [17] M. Formentin, A. Lovison, A. Maritan, and G. Zanzotto, Phys. Rev. E **90**, 012817 (2014).
- [18] P. Wang, T. Lei, C. H. Yeung, and B.-H. Wang, EPL (Europhysics Letters) **94**, 18005 (2011).
- [19] T. Zhou, Z.-D. Zhao, Z. Yang, and C. Zhou, EPL (Europhysics Letters) **97**, 18006 (2012).
- [20] E. R. Colman and D. Vukadinovic Greetham, Phys. Rev. E **92**, 012817 (2015).
- [21] M. Karsai, K. Kaski, A.-L. Barabasi, and J. Kertesz, Sci. Rep. **2**, 397 (2012).
- [22] H.-H. Jo, J. I. Perotti, K. Kaski, and J. Kertesz, Phys. Rev. E **92**, 022814 (2015).
- [23] H.-H. Jo, M. Karsai, J. Kertesz, and K. Kaski, New Journal of Physics **14**, 013055 (2012).
- [24] J. Kim, D. Lee, and B. Kahng, PLoS ONE **8**, e58292 (2013).
- [25] J. L. Iribarren and E. Moro, Phys. Rev. Lett. **103**, 038702 (2009).
- [26] R. Lambiotte, L. Tabourier, and J.-C. Delvenne, The European Physical Journal B **86** (2013).
- [27] C. Wang and B. A. Huberman, Sci. Rep. **2**, 633 (2012).
- [28] M. Formentin, A. Lovison, A. Maritan, and G. Zanzotto, ArXiv e-prints (2014), arXiv:1405.5726 [physics.soc-ph].
- [29] E. Ferrara, R. Interdonato, and A. Tagarelli, Proceedings of the 25th ACM Conference on Hypertext and Social Media , 24 (2014).
- [30] J. Ratkiewicz, S. Fortunato, A. Flammini, F. Menczer, and A. Vespignani, Phys. Rev. Lett. **105**, 158701 (2010).
- [31] M. Coscia, International AAAI Conference on Weblogs and Social Media (2013).
- [32] S. A. Myers and J. Leskovec, Proceedings of the 23rd International Conference on World Wide Web , 913 (2014).
- [33] Algorithms and Models for the Web Graph **8882**, 132 (2014).
- [34] Q. Zhao, M. A. Erdogdu, H. Y. He, A. Rajaraman, and J. Leskovec, KDD15 (2015).
- [35] D. Boyd, S. Golder, and G. Lotan, System Sciences (HICSS), 2010 43rd Hawaii International Conference on , 1 (2010).
- [36] T. R. Zaman, R. Herbrich, J. V. Gael, and D. Stern, Computational Social Science and the Wisdom of Crowds Workshop (colocated with NIPS 2010) (2010).
- [37] D. V. Greetham and Ward, Proceedings of the 2nd International Workshop on Dynamic Networks and Knowledge Discovery , 73 (2014).
- [38] M. D. Domenico, A. Lima, P. Mougel, and M. Musolesi, Sci. Rep. **3**, 2980 (2013).
- [39] S. Shinomoto, K. Shima, and J. Tanji, Neural Comput. **15**, 2823 (2003).
- [40] K. Miura, M. Okada, and S. ichi Amari, Neural Comput. **18**, 2359 (2006).
- [41] T. Omi and S. Shinomoto, Neural Comput. **23**, 3125 (2011).
- [42] S. Shinomoto, H. Kim, T. Shimokawa, N. Matsuno, S. Funahashi, K. Shima, I. Fujita, H. Tamura, T. Doi, K. Kawano, N. Inaba, K. Fukushima, S. Kurkin, K. Kurata, M. Taira, K.-I. Tsutsui, H. Komatsu, T. Ogawa, K. Koida, J. Tanji, and K. Toyama, PLoS Comput Biol , e1000433 (2009).
- [43] C. Sanli and R. Lambiotte, PLoS ONE **10**, e0131704 (2015).
- [44] C. Sanli and R. Lambiotte, Modeling and Mining Temporal Interactions: Papers from the 2015 ICWSM Workshop , 8 (2015).
- [45] J. Leskovec and A. Krevl, SNAP Datasets: Stanford Large Network Dataset Collection (2014).
- [46] J. Saramaki, E. A. Leicht, E. Lopez, S. G. B. Roberts, F. Reed-Tsochas, and R. I. M. Dunbar, Proceedings of the National Academy of Sciences **111**, 942 (2014).
- [47] P. d. Meo, E. Ferrara, F. Abel, L. Aroyo, and G.-J. Houben, ACM Trans. Intell. Syst. Technol. **5**, 14 (2014).
- [48] B. A. Huberman, D. M. Romero, and F. Wu, First Monday **14** (2009).
- [49] B. Goncalves, N. Perra, and A. Vespignani, PLoS ONE **6**, e22656 (2011).
- [50] L. Weng, A. Flammini, A. Vespignani, and F. Menczer, Sci. Rep. **2**, 335 (2012).
- [51] J. P. Gleeson, J. A. Ward, K. P. O'Sullivan, and W. T. Lee, Phys. Rev. Lett. **112**, 048701 (2014).
- [52] U. Cetin and H. O. Bingol, Phys. Rev. E **90**, 032801 (2014).
- [53] J. Borge-Holthoefer, R. A. Banos, S. Gonzalez-Bailon, and Y. Moreno, Journal of Complex Networks (2013).
- [54] H.-H. Jo, R. K. Pan, and K. Kaski, Phys. Rev. E **85**, 066101 (2012).